



The memorization of in-line sensorimotor invariants: toward behavioral ontogeny and enactive agents

Pierre de Loor, Kristen Manac'H, Pierre Chevaillier

► To cite this version:

Pierre de Loor, Kristen Manac'H, Pierre Chevaillier. The memorization of in-line sensorimotor invariants: toward behavioral ontogeny and enactive agents. *Artificial Life and Robotics*, 2014, 19 (2), pp.127-135. 10.1007/s10015-014-0143-3 . hal-01066401

HAL Id: hal-01066401

<https://hal.science/hal-01066401>

Submitted on 21 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pierre De Loor, Kristen Manac'h and Pierre Chevaillier
Lab-STICC – UEB – ENIB, F29200 Brest, France

The Memorization of In-Line Sensorimotor Invariants : Toward Behavioral Ontogeny and Enactive Agents

Abstract This paper presents a behavioral ontogeny for artificial agents based on the interactive memorization of sensorimotor invariants. The agents are controlled by Continuous Timed Recurrent Neural Networks (CTRNNs) which bind their sensors and motors within a dynamic system. The behavioral ontogenesis is based on a phylogenetic approach: memorization occurs during the agent's lifetime and an evolutionary algorithm discovers CTRNN parameters. This shows that sensorimotor invariants can be durably modified through interaction with a guiding agent. After this phase has finished, agents are able to adopt new sensorimotor invariants relative to the environment with no further guidance. We obtained these kinds of behaviors for CTRNNs with 3 to 6 units, and this paper examines the functioning of those CTRNNs. For instance, they are able to internally simulate guidance when it is externally absent, in line with theories of simulation in neuroscience and the enactive field of cognitive science.

Key words Ontogeny, Evolutionary robotics, Embodied agent, Sensorimotor invariants, CTRNN, Memory, Enaction.

Introduction

Embodied cognitive sciences theories claim that sensorimotor invariance is of key importance – even more so than representation – for the constitution of cognition [1],[2]. Sensorimotor invariance is the systematic coupling of sensory and motor patterns: *it specifies the structure of the rules governing the sensory changes produced by various motor skills* [3]. Accordingly, cognitive processes, such as reasoning, learning, memorization, and categorization, emerge from the interactions between dynamic components operating at different time-scales [4]. In cognitive psychology, several au-

thors have conceptualized human interactions with dynamic systems [5], and Beer's seminal work in cognitive science was inspired by dynamic approaches [6]. This theoretical position has led to the development of models that produce cognitive behaviors without internal representations of the external world. The advantages of the dynamic approach include its ability to deal with novelty, continuous timed, variability and nondeterministic characteristics of cognition. For example, Iizuka and Di Paolo showed that agents were able to exhibit behavioral preference when exposed to two stimuli [7]. When both stimuli were presented simultaneously, the agents were still able to develop a behavioral preference since they adopted one of the evolved behaviors. The preference could change, conditioned on, but not determined by, the environment. Continuous Timed Recurrent Neural Networks (CTRNNs) can represent these kinds of models in a fully integrated form that is non-modular and does not utilize a learning-specific mechanism. Thanks to evolutionary approaches, they can be configured to exhibit multiple cognitive abilities. For example, Wood and Di Paolo used CTRNNs to simulate the famous 'A not B' task proposed by Piaget in his epistemological studies of young children [8], [9]. Tuci *et al.* used evolved agents controlled by dynamic neural networks to conduct experiments on associative learning [10]. By using synaptic plasticity, these agents were able to associate a landmark with a goal and to memorize how the goal was related to the landmark. In a study on integrated associative learning in dynamic neural networks, evolved agents successfully associated a motor command, evaluated in a binary form, with a stimulus whose value laied on a continuum [11].

One remaining challenge is to create agents able to behaviorally evolve *during their lifetime* according to their interactions with the environment. This idea is tied to the work of Lindblom and Ziemke on robot ontogenesis in social situations [13], which is in turn based on Vygotsky's and Piaget's

Preprint of : De Loor, P.; Manac'h, K. & Chevaillier, P. (2014), 'The memorization of in-line sensorimotor invariants: toward behavioral ontogeny and enactive agents', *Artificial Life and Robotics* 19(2), 127-135.

DOI: <http://dx.doi.org/10.1007/s10015-014-0143-3>

research. Psychologists have showed that the behavioral development of a child coincides with the development of her sensorimotor, social, and reasoning skills [12]. More fundamentally, behavioral ontogeny results from a process involving not only the gradually changing individual, but also the environment where she acts. It implies an active role of the environment which reacts differently depending on the behavior of the agent. The understanding and modeling of this ontological process is far from settled.

It is still necessary to perform preliminary interactive online shaping of any behaviors in order to obtain behavioral ontogeny. Crucially, such behaviors must be memorized, and depend on the perceived environment of the agent. Since sensorimotor invariants are the basic components of cognitive capabilities, it must be possible to maintain any behavior from the interaction history to ensure the creation of ontogenetic agents. This implies support for the general memorization of sensorimotor invariants, which this paper addresses in the following sections.

Section 2 uses the enactive stance to explain the differences between the classical concept of memorization and sensorimotor invariants memorization. Section 3 proposes how evolutionary robotics might use this capability to create autonomous agents. Section 4 presents results obtained from our experiments in this area, and identifies common behavioral and structural properties of the resulting CTRNNs. Section 5 draws conclusions and discusses the next stage of our work.

Dynamic approach of memory

Traditionally, memory is represented by a variable which can be assigned to a value. For example, when an agent perceives a target, she sets a variable called “target” to the value “true”. Such explicit representations are not well-suited to the dynamic approach. For example, some evolutionary robotics researches utilize behaviors that apply the principles of memorization to CTRNNs, or other recurrent networks, without representations [13]. However, these systems still view memory in a classical sense and face the problem of generalization: if a solution exists, then it is impossible for the agent to adopt another behavior, especially one different from its initial form. Unfortunately, these approaches re-introduce representations into the reactive model and thus faced by the same problem as ordinary *representationalist approaches*: they fail to introduce novelty at the ontogenetic level. Our first proposition overcomes this problem by rethinking the notion of memory in light of the enactive perspective. Agents do not memorize a perceived object but adopt a habit by memorizing behaviors, which are basically sensorimotor invariants resulting from interactions with the environment. When agents are able to memorize behavior, they have a gen-

eral memorization ability limited only by their sensorimotor features, as is the case for living beings. This technique implies more complex principles than just the utilization of a variable to perceive and make decisions. Memory, or habit, is shaped by the environment, and can potentially vary or evolve in infinite ways. To model this, we must introduce ‘active’ interaction between the agent and its environment. The active role of the environment in the cognitive process brings it into line with the enactive stance, and our work uses this relationship for guidance.

Figure 1 shows a scenario that exemplifies our approach. It has three main elements: (i) L , a set of landmark $\{L_1, L_2\}$, (ii) an agent A , called the ontogenetic agent, which perceives the landmarks in L , and (iii) a guide B ; B interacts with A to change its L -related behavior. During phase 1, agent A has an L -related behavior that depends on its distances from the landmarks, as calculated by its sensors. This behavior is implemented as a function f , of A and L . Agent A also has a “social” behavior, a function called g , utilizing both A and B . During phase 2, guide B observes A and interacts with her by way of a continuous signal S_B which changes f into a new behavior f' . During this phase, agent A combines the f and g functions which means there is a progressive transformation of f into f' . Eventually, guide B stops interacting with A , and A maintains its new sensorimotor behavior f' .

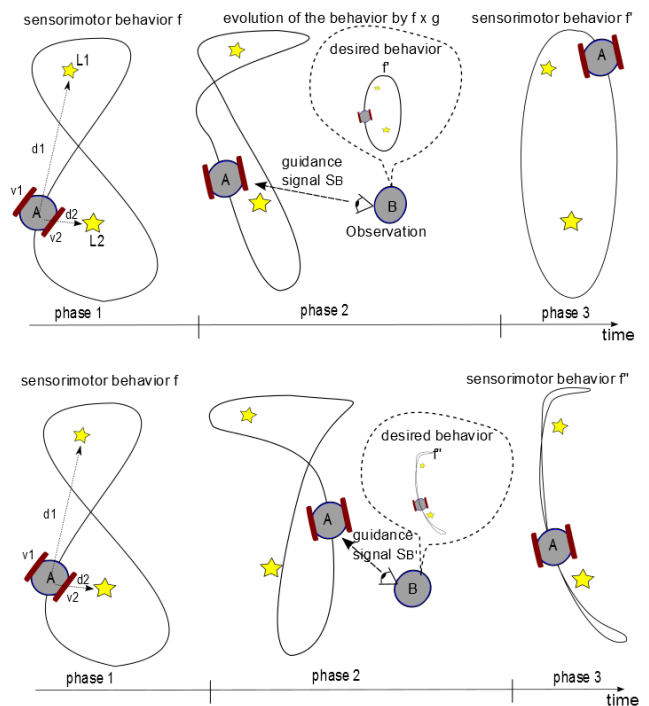


Fig. 1 Illustrations of behavioral evolution based on guidance. During phase 1, agent A adopts a sensorimotor invariant relative to L_1 and L_2 . This is achieved by the association between sensor's values d_1 , d_2 (distance between the agent and L_1 and L_2 respectively) and actuator's values, v_1 and v_2 (velocities of the two wheels of the agent). During phase 2, B guides agent A via an in-line signal (S_B). Agent A responds by progressively modi-

fying its behavior. B “knows” what new behavior it wants for A (it is drawn as a dotted speech balloon above B in the figure). The guidance signal is actively generated according to A's behavior as observed by B. After phase 2, B ends its guidance, and A maintains its new sensorimotor invariant relative to L1 and L2.

The consequence is that A memorizes a new sensorimotor behavior f' and forgets the previous behavior f . The generalization of this approach requires the reproduction of this evolution with different kinds of behaviors, as illustrated at the bottom of Figure 1 where another guidance signal S_B changes the behavior f into f'' , different from f' . These principles differ from traditional reinforcement learning [14] on two points: 1) no explicit internal agent's states are used, so the behavior is not constrained by discrete symbols, and could theoretically develop any level of complexity; 2) guidance is a continuous signal relative to the actual behavior of the agent (it is neither discrete nor a reward handed out following the agent's success). The evolution of this signal increases if the agent deviates from the desired behavior and decreases if the agent tends towards good behavior, which is exploited by the internal dynamics of the agent to correct its behavior. These points are crucial for addressing the notion of behavioral ontogeny. It is possible to extract three basic principles:

- 1) The agent must be able to adopt sensorimotor invariants relative to elements perceived in the environment (L) without an explicit representation of that environment. Ideally, the number of behaviors must be unlimited (f , f' , f'' ...).
- 2) The agent interacts with a guide (B) by means of a continuous signal with an implicit meaning (S_B). The agent combines its behavior relative to the environment and its behavior relative to the guide to progressively transform its environmental behavior.
- 3) The agent is able to maintain its new behavior after the period of guidance has finished.

Our system utilizes these principles, applied to a very simple environment, so that the analysis and complexity of the evolutionary approaches are tractable.

Experimental Setup

The experimental environment has a single dimension and one landmark $L=\{L_1\}$. So it is more simple than the illustration of the figure 1 but still preserves its principles. The agent can perceive its position relative to the landmark, and the landmark can move with varying speed and direction based on random parameters (at each step of the simulation, a random number in $[-0.005, 0.005]$ is added to the speed of the landmark and if the position of the landmark reaches the limits $+5$ or -5 , this speed is inversed). The initial L-related behavioral invariant f for the agent A maintains a specified desired position d from the landmark irrespective of its movements. It is the phase 1 of the figure 1. The evolution towards a L-related behavior f' changes the desired position from d to d' . The capability is generalized by allowing the

time of the change and the new position d' to be randomly set.

The agent A is controlled by a fully connected CTRNN of 3 to 6 units. The behavior of each unit i of the network is governed by the differential equation:

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^k w_{ij} \sigma[y_j + b_j] + I_i$$

y_i is the activation of the i^{th} unit, b_j is a bias, and $\sigma(x) = 1/(1 + \exp(-x))$ is a sigmoid activation function. w_{ij} is the weight of the connection from the j^{th} to the i^{th} unit, I_i is an input unit, and k is the number of units in the network. In our system, the position of the agent relative to the landmark L_1 is sent to one unit, and the guidance signal of B sent to another unit. Likewise, the output of one unit determines the speed of the agent.

Each time that a new behavioral invariant should be adopted by the agent A, the guidance signal $S_B(t)$ is sent out for a period of 250 time units (this is termed phase 2 on figure 1). The signal is continuous and modulated by the agent's trend towards its new behavior. More specifically, if $d_{LA}(t)$ is the position of the agent relative to the landmark at time t and $d_{\text{new}}(t)$ is the new position that should be adopted, then the trend toward the new behavior is defined as:

$$\Delta(t) = \frac{d(|d_{\text{new}}(t) - d_{LA}(t)|)}{dt}$$

Then, if the position of the agent approaches the desired position, $\Delta(t)$ becomes negative.

The guidance function is:

$$S_B(t) \begin{cases} S_B(t-1) * (1 + \Delta(t)) & \text{if } (\Delta(t) < 0) \\ S_B(t) = (d_{\text{new}}(t) - d_{LA}(t)) & \text{else} \end{cases}$$

Hence, if the agent's behavior is heading towards the expected behavior, the guidance signal will decrease rapidly. Conversely, an increasing distance between the current behavior and the desired behavior will make the signal rise. Figures 2, 6, and 7 give an idea of the evolution of the signal relative to the behavior of the agent.

Guidance is only sent for 250 time units, but is computed throughout the experiment in order to judge the agent's fitness. This fitness measure is inversely proportional to the need for guidance even when it is inactive (during phase 3).

The genetic technique employed is Harvey's microbial algorithm which is well-suited for the optimization of CTRNNs [15]. It utilizes the principle of tournament to guaranty the preservation of potential niches, and performs transfusion and mutation on real numbers with a Gaussian vector mutation.

The variables involved are τ and b for each unit, and the weights w_{ij} of each connection.

The fitness of an agent is defined over the course of 21000 time units, and the mean number of behavior invariants to memorize during this time is $\overline{M}=20$. No exact value is given for this number because it depends on the randomly fixed instants at which a new distance $d_{new}(t)$ is selected. If $K(i)$ is the length of the i^{th} memorization episode, then the agent's fitness for that episode is:

$$fitness_i = 1 - \sum_{t=250}^{t=K(i)} \left\{ \frac{|d_{new}(t) - d_{LA}(t)|}{20 * (K(i) - 250)} \right\}$$

$d_{new} \in [-10, +10]$ is positive when the agent is in front of the landmark, and negative when it is behind it. In a similar way, $d_{LA} \in [-10, +10]$. The fitness of an agent who doesn't need guidance after 250 time units is maximal, and equals 1. Conversely, the fitness declines if an agent needs guidance after 250 time units. The agent's global fitness is the mean of its fitness over M episodes:

$$fitness = \frac{\sum_{i=0}^{i=M} \{fitness_i\}}{M + 1}$$

Memorization criteria

In addition to the fitness of the evolutionary steps, we define two criteria for the detailed evaluation of agent behavior after it has memorized a sensorimotor invariant. The first criterion considers the *precision* of the memory since memory is a behavior, not a value. It is observable through continuous variables, and should be fairly close to the desired behavior. Consequently, *memoryPrecision* evaluates the gap between the theoretical position of the agent (d_{new}) and her real position (d_{LA}).

The agent's ability to memorize the correct sensorimotor invariant is assessed by a repeated test scenario. After the first memorization phase, the adopted sensorimotor invariant is randomly modified to become d_{new} . The agent is guided over a period of 250 time units and left unguided for a test period T_{test} of 3000 time units. This test is repeated 20 times for each agent:

$$memoryPrecision = \frac{1}{20} \left[\sum_{i=1}^{20} \frac{1}{T_{test}} \left[\sum_{t=i}^{T_{test}} |d_{new}^i - d_{LA}(t)| \right] \right]$$

If the actual behavior of the agent is tending towards the desired behavior, then *memoryPrecision* will approach 0. How-

ever, if the value is rising, then the new behavior is deviating from the desired one.

The second criterion (*memoryDuration*) evaluates the time after which the memorization is considered to be lost. After the first memorization phase, a new sensorimotor invariant (d_{new}) is randomly fixed. The agent is guided for 250 time units and then released from the guidance. The *memoryDuration* value will be the time at which the difference $d_{new} - d_{LA}$ exceeds a threshold distance $d > 1$ (which corresponds to 5% of the maximum difference between the two distances). This test is performed 20 times per agent.

Results

Figure 2 shows the typical agent behavior resulting from an evolutionary run. The grey regions correspond to guidance periods of 250 time units. The guidance signal is shown at the top, and the behavior of the landmark L and agent appear at the bottom, in red and green respectively. The adoption distance is depicted in the middle graph of Figure 2. The landmark movement evolves randomly to ensure that the agent uses the landmark as an information source rather than following a predefined trajectory. More precisely, the position of the landmark (p_{LA}) is computed from a variable s_{LA} which represents a movement of the landmark:

$$p_{LA}(t) = p_{LA}(t-1) + s_{LA}(t).$$

If $s_{LA}(t) < 0$ the landmark moves toward the bottom. Else, it moves towards the top. To maintain an observable behavior, if $|p_{LA}(t)|$ becomes greater than 5, $s_{LA}(t)$ is inversed. The randomization is on $s_{LA}(t)$ is as follows:

$$s_{LA}(t+1) = s_{LA}(t) + r(t)$$

where $r(t)$ is a random number between -0.005 and 0.005. By this way, the movement of the landmark is not really noisy but can vary in an infinite manner. It follows a trajectory which evolves randomly during time (see Figure 2 for instance).

Guidance stops at the end of each guidance period (shown in grey), and the agent moves freely for a random time. The results show that the agent preserves its distance from the landmark, indicating that the sensorimotor invariant has been memorized. When a new distance is randomly defined, guidance restarts. Figure 2 illustrates that, independent of the distance and movement of the landmark, the agent is able to memorize the sensorimotor invariant.

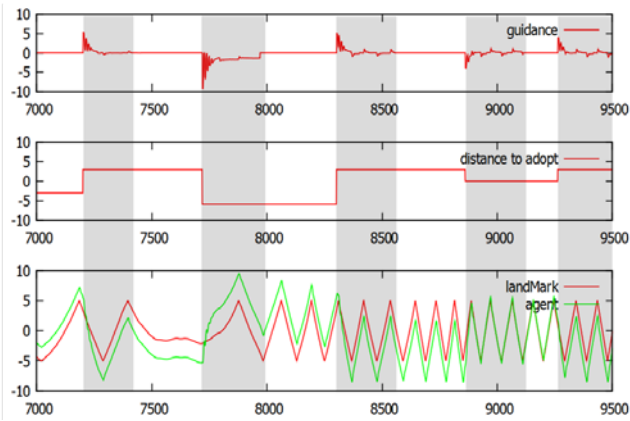


Fig. 2 The memorization of sensorimotor invariants. Each shaded area is a guidance period during which the agent is helped to maintain a specific distance (randomly chosen) from the landmark. Each white area is a random duration when the agent is unguided.

We varied the number of units in the CTRNN from 3 to 6 to evaluate its impact on the different criteria. Actually, we did not use genetic algorithms able to add and remove neurons or connections, as used by [16] for instance. There were two motivations for that: First we tried this solution and the time consumed by the algorithms to test different network configurations was prohibitive. The second was that we needed to master the number of units in order to observe the differences and common points in terms of structure and behaviors provided by different structures of networks. For example, we need to know if a small number of units can lead to an acceptable behavior and if this behavior could be improved by a higher number of units. Genetic algorithms were run 40 times for each configuration and stopped after 10000 generations. Figure 3 shows a box plot of the fitness for each configuration. The fitness means are 0.58 (for 3 units), 0.70 (for 4 units), 0.743 (for 5 units), and 0.736 (for 6 units). The differences between these values are only significant for the 3 units agent. The Welch fitness test between the 3 and 4 unit agents lets us reject the null hypothesis with a p -value = 0.0021. But we cannot reject it when comparing 4 and 5 units, or 5 and 6 units, which gave p -values of 0.059 and 0.61 respectively. We obtained similar results when considering *memoryPrecision* and *memoryDuration*. However, as Figure 3 shows, the best fitness does not necessary correspond to the best memory precision or duration. Fitness tends to minimize guidance while the genetic algorithm selects agents that are very sensitive to that guidance. This sensitivity ensures that the agent rapidly reaches the desired behavior but does not mean that it will adopt that behavior over the long term. Nevertheless, it is possible to find agents which combine good fitness (i.e. with a value close to 1), good memory duration (greater than 1000) and memory precision (less than 3) for each of the four configurations (3,4,5 and 6 units per agent).

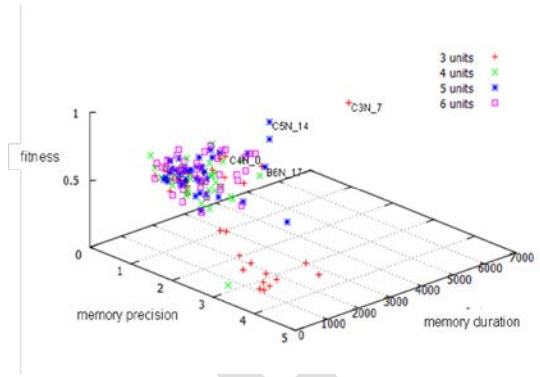


Fig. 3 Fitness, memory precision, and memory duration for the best agents after 160 runs of the genetic algorithm (40 runs per configuration).

To study the behavior of good agents, we created a CTRNNtest set containing agents with good overall performance across all the categories (see Table 1).

Id	nb units	fitness	memory precision	memory duration
C3N_7	3	0.78	1.66	6058
C4N_0	4	0.73	1.59	2059
C5N_14	5	0.84	1.67	3501
B6N_17	6	0.66	2.17	2100

Table 1 : The CTRNNtest set of CTRNNs which exhibit good properties for behavioral memorization by guidance.

Interpretation

We explored the evolutionary approach using two techniques: an enriched graphical representation of each network in the CTRNNtest set (see Figure 4), and the examination of the time evolution of the system during and after memorizing (see Figures 6 and 7). Figure 4 lets us compare the values of each arc weight and each unit time constant because the thickness of the arrows between two units i and j is proportional to the weight w_{ij} , and the lightness of a node j is related to the τ_j value. Despite each network being the result of totally independent runs of the genetic algorithm, they still share several common features:

- 1) The unit which reads the guidance signal (unit 2) has a high self-loop weight.
- 2) There is a high weight from unit 2 towards the unit which reads the distance between the landmark and the agent (d_{LA}) (unit 1).
- 3) The unit which reads the distance d_{LA} has a weak self-loop weight (unit 1).
- 4) The unit which reads the guidance signal (unit 2) has a higher time constant than the unit which reads the distance d_{LA} (unit 1).
- 5) There is a unit with a weak time constant and a high self-loop weight located between the unit which reads the guidance signal (unit 2) and outputs the *Speed* signal.

These five points lead to the following interpretation for our approach. Guidance is maintained for some time thanks to

the point 1, and mainly processed by the node that evaluates the d_{LA} distance. Guidance has low direct impact on the node controlling the speed (points 2 and 5), which allows the guidance signal to decrease without changing the agent distance from the landmark. The previous d_{LA} value is not retained (point 3), which permits greater reactivity when the distance does alter. Agent speed is very self-dependent (point 5), except in the case of the C4N_0 network where a self-loop is

performed on unit 4. Speed self-dependency highly influences the unit which evaluates landmark distance. As a consequence, an agent can autonomously maintain its speed while still being capable of reacting rapidly to a d_{LA} change, especially when its speed is high.

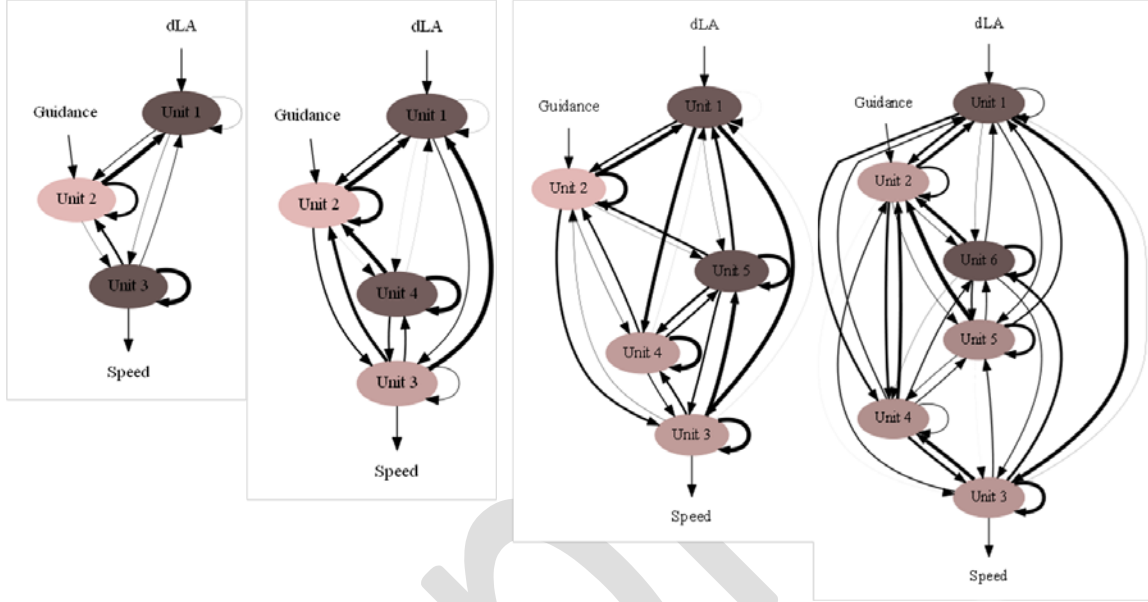


Fig. 4 The four networks in the CTRNNtest set. The thickness of an arc between unit i and unit j corresponds to w_{ij} . The darkness of a unit j is inversely proportional to τ_j .

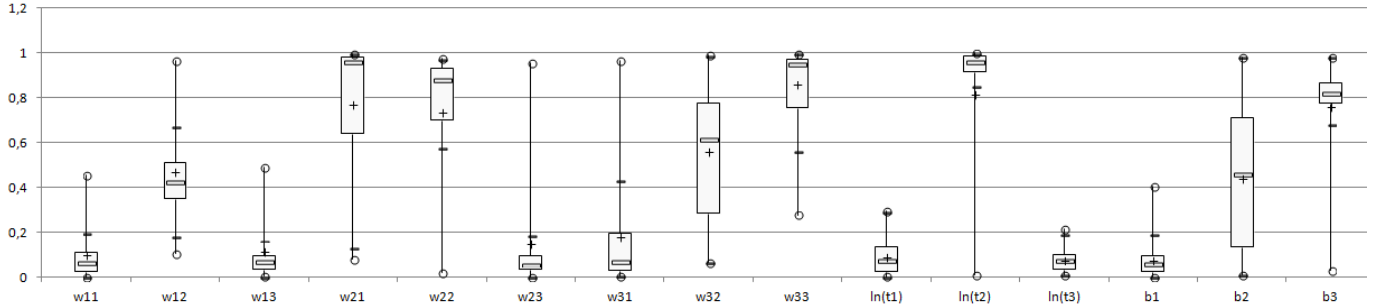


Fig 5: Boxplot for the parameters (w_{ij} , b_i , $\ln(\tau_i)$) of 40 networks of 3 units resulting from two runs of the genetic algorithm during 10000 generations. “□” is the median, the bounds of the box represent the first and third quartiles, “+” is the mean, “-” are the lower and upper limits and “o” are the min and max values.

Figure 5 allows the comparison between 40 networks of 3 units resulting of the genetic algorithm. It confirms that the common features are globally shared by each network (the interval between the first and the third quartiles is often short) : For instance, the high value of τ_2 compared to τ_1 and τ_3 ($\ln(\tau_2) \approx 13,45 * \ln(\tau_1) \approx 13,5 * \ln(\tau_1)$) or the high value of w_{22} and w_{33} compared to w_{11} ($w_{11} \approx w_{22} * 0.069 \approx w_{33} * 0.064$). However, in rare cases (see the min or max values of the parameters), some networks present a totally different configuration.

It confirms that there is not only one way to reach the memorization objectives, even if the previous common features are dominant. Similar conclusions can be made for 4 to 6 units networks.

Figure 6 shows the overall behavior of the C3N_7 network. The graph at the bottom of the figure is a time series for the activation of its units while the top graph represents the evolution of the guidance signal, and the movements of the landmark and agent. Before time $t=850$, the agent maintains its initial sensorimotor invariant (a distance $d_{LA} = d1$ from the

landmark), but at $t=850$ a new distance d_2 is adopted. The guidance signal is authorized during the guidance period and the distance d_{LA} approaches the value d_2 . The figure illustrates how the activation of units 1 and 3 follow the guidance oscillations.

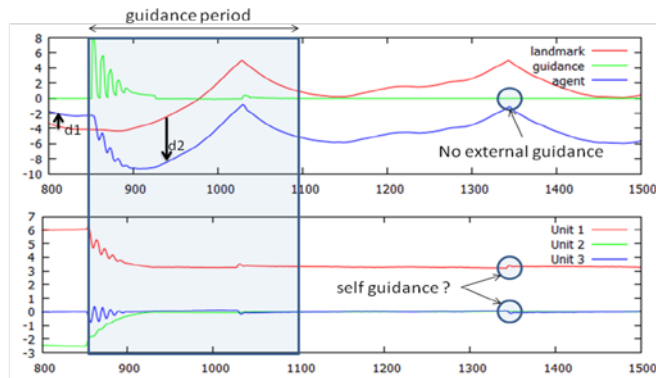


Fig. 6 Behavior of the CTRNN C3N_7 network.

The guidance signal stops at the end of the guidance period, and the agent maintains a distance d_2 from the landmark even while it is moving. At time 1350, the landmark's direction of movement reverses which causes a small change in the distance d_{LA} . Since guidance has stopped by this time, no guidance signal is sent to the agent. Despite this, the agent reverses its movement because it has memorized the new sensorimotor invariant. An interesting phenomena occurs at time $t=1350$, when units 1 and 3 produce small variations similar to those observed when guidance was authorized. The dynamic system is executing as if it was being guided, which suggest that this system is able to simulate guidance for itself. We observed this phenomenon for all agents resulting from the selection process.

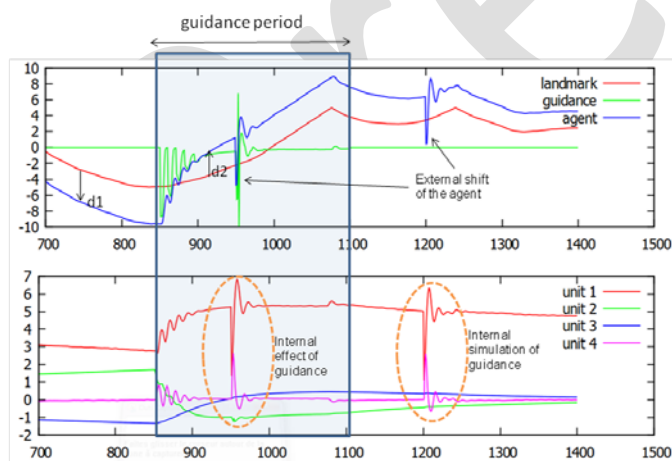


Fig. 7 : Behavior of the CTRNN C4N_0 network. At time 850, the sensorimotor distance changes from d_1 to d_2 . At time 950, the agent is externally shifted and the guidance signal progressively corrects the distance between the agent and the landmark, and units 1 and 4 oscillate. At time 1200, the position of the agent is shifted again when there is no guidance signal but the CTRNN simulates its own guidance to find a good sensorimotor invariant.

To confirm these phenomena, we experimented with a more extreme situation, by rapidly modifying the agent's position twice in the C4N_0 network (as shown in Figure 7). The first move occurred during the guidance period (at time 950) and the second after its finish (at time 1200). The results show that the units which were influenced by the guidance signal during the guidance period are able to simulate this influence even when there is no guidance. It is remarkable how this internal simulation is related to the previous guidance signal because it suggests that the agent is able to "retain" the necessary internal simulation based on the previous guidance phase. This shows that interactions leave their mark on the dynamic of the system. Similar observations were made for all the networks in CTRNNtest.

Conclusions

In the field of evolutionary robotics, memorization is generally addressed using a 'traditional' approach, i.e., the perception of a discrete stimulus that governs the selection of behaviors. In this paper, we proposed another kind of memorization which follows the enactive stance by having interaction that modulates the agent's sensorimotor invariants. This modulation occurs *during the agent lifetime*, i.e. at the ontogenesis level. The genetic algorithm does not find an agent able to distinguish one stimulus from another, but rather one able to learn various in-line sensorimotor invariants thanks to its interaction with a guide. We argue that this result is a step toward ontogenetic behavior. Indeed, as sensorimotor invariants are the basis of all cognitive ability, sensorimotor invariant memorization is a crucial point for enactive cognition. Currently, the generated sensorimotor behaviors are simple because of the agent's poor action and perception capabilities (i.e. one sensor, one actuator, and one guidance signal), but we plan to examine more complex sensorimotor configurations. The main problem is that the genetic algorithm is rapidly faced by a huge search space, and heuristics must be discovered that reduce its size. This seems possible because all CTRNNs share common structural and dynamic properties, which might become predefined core components of more complex CTRNNs in the future. Also since a guide interacts with agents online, we can study more complex behaviors by replacing the guide by a human who could utilize higher levels of cognition [17]. In our experiments, we observed that all CTRNNs are able to simulate the guidance signal internally, which opens up interesting prospects for a link between ontogenetic agents and the simulation theory of neuroscience [18].

References

- [1] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind*, MIT Press. 1991.
- [2] L. Shapiro, "Travelling in Style from Standard Cognitive Science to Embodied Cognition," *ConstruTivist*, pp. 231–233, 2011.
- [3] A. Noë, *Action in Perception*. Cambridge, MA: MIT Press, 2004.
- [4] L. Smith and E. Thelen, "Development as a dynamic system," *Trends Cogn. Sci.*, vol. 7, no. 8, pp. 343–348, 2003.

- [5] W. H. Warren, "The dynamics of perception and action.," *Psychol. Rev.*, vol. 113, no. 2, pp. 358–89, Apr. 2006.
- [6] R. D. Beer, "Dynamical approaches to cognitive science," *Trends Cogn. Sci.*, vol. 4, no. 3, pp. 91–99, 2000.
- [7] E. Di Paolo and H. Iizuka, "How (Not) to Model Autonomous Behavior," *BioSystems*, vol. 91, no. 2, pp. 409–423, 2008.
- [8] R. Wood and E. A. Di Paolo, "New models for old questions: Evolutionary robotics and the 'A not B' error," in *Proceedings of the 9th European Conference on Artificial life ECAL 2007.*, 2007.
- [9] J. Piaget, *Play, Dreams and Imitation in Children*. 1951.
- [10] E. Tuci, M. Quinn, and I. Harvey, "An evolutionary ecological approach to the study of learning using a robot based model," 2003.
- [11] E. Izquierdo, I. Harvey, and R. Beer, "Associative Learning on a Continuum in Evolved Dynamical Neural Network," *Adapt. Behav.*, vol. 16, no. 6, pp. 361–384, 2008.
- [12] G. Kugiumutzakis, "Genesis and development of early infant mimesis to facial and vocal models," in *Imitation in Infancy*, J. Nadel and G. Butterworth, Ed. Cambridge university Press, 1999, pp. 36–59.
- [13] O. Charles, P. Tony, and D. St, "With a little help from selection pressures : evolution of memory in robot controllers," 2009.
- [14] R. Sutton and A. Barto, *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.
- [15] I. Harvey, "The microbial genetic algorithm," *Adv. Artif. Life. Darwin Meets von Neumann*, pp. 126–133, 2011.
- [16] J. B. Mouret and S. Doncieux, "Using behavioral exploration objectives to solve deceptive problems in neuro-evolution," in *Proceedings of the 11th Annual conference on Genetic and evolutionary computation, GECCO'09*, 2009, pp. 627–634.
- [17] P. De Loor, K. Manac'h, and J. Tisseau, "Enaction-Based Artificial Intelligence: Toward Co-evolution with Humans in the Loop," *Minds Mach.*, vol. 19, no. 3, pp. 319–343, Oct. 2009.
- [18] S. H. E. Raos Vassilis, Evangeliou Mina N., "Mental Simulation of Action in the Service of Action Perception," *J. Neurosci.*, vol. 27, no. 46, pp. 12675–12683, 2007.